

ARBRES DE DECISION ET FORETS ALEATOIRES.

Philippe Courcoux, Evelyne Vigneau

StatSC Oniris
Nantes / France
philippe.courcoux@oniris-nantes.fr

Résumé

Les arbres de décision et les forêts aléatoires font partie des techniques d'apprentissage statistique visant à la prédiction dans le cadre de la régression ou de la discrimination. Développées par Breiman, ces méthodes ont fait l'objet de développements nombreux et sont accessibles maintenant dans la plupart des packages statistiques.

Le workshop vise à présenter les arbres de décision de type CART (Classification And Regression Trees) ainsi que les forêts aléatoires et leurs applications potentielles dans le cadre du traitement de données sensorielles.

Les points abordés seront :

- Arbres CART (construction et élagage dans le cadre de la régression et de la discrimination)
- Forêts aléatoires (méthode, importance des variables, prédiction)
- Stratégies de sélection de variables
- Variantes et développements (arbres de décision multivariés, boosted regression trees...)
- Présentation de packages R pour la mise en œuvre des méthodes présentées.

Les techniques d'arbres de décision et de forêts aléatoires seront illustrées par des cas d'étude dans le cadre du traitement de données sensorielles. On présentera des applications à la modélisation de données de profil en fonction de variables instrumentales et à la mise en relation de données de préférence et de profil sensoriel.

Keywords - Decision trees, CART, Random Forests, Machine Learning.

REFERENCES

- [1] Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). Classification and Regression Trees. New-York, Chapman & Hall. Brown B, Aaron M (2001) In: Smith J (ed) The rise of modern genomics, 3rd edn. Wiley, New York
- [2] Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32.
- [3] Genuer, R., Poggi, J.-M. and Tuleau-Malot, C. (2010). Variable Selection using Random Forests. Pattern Recognition Letters, 31, 2225-2236.